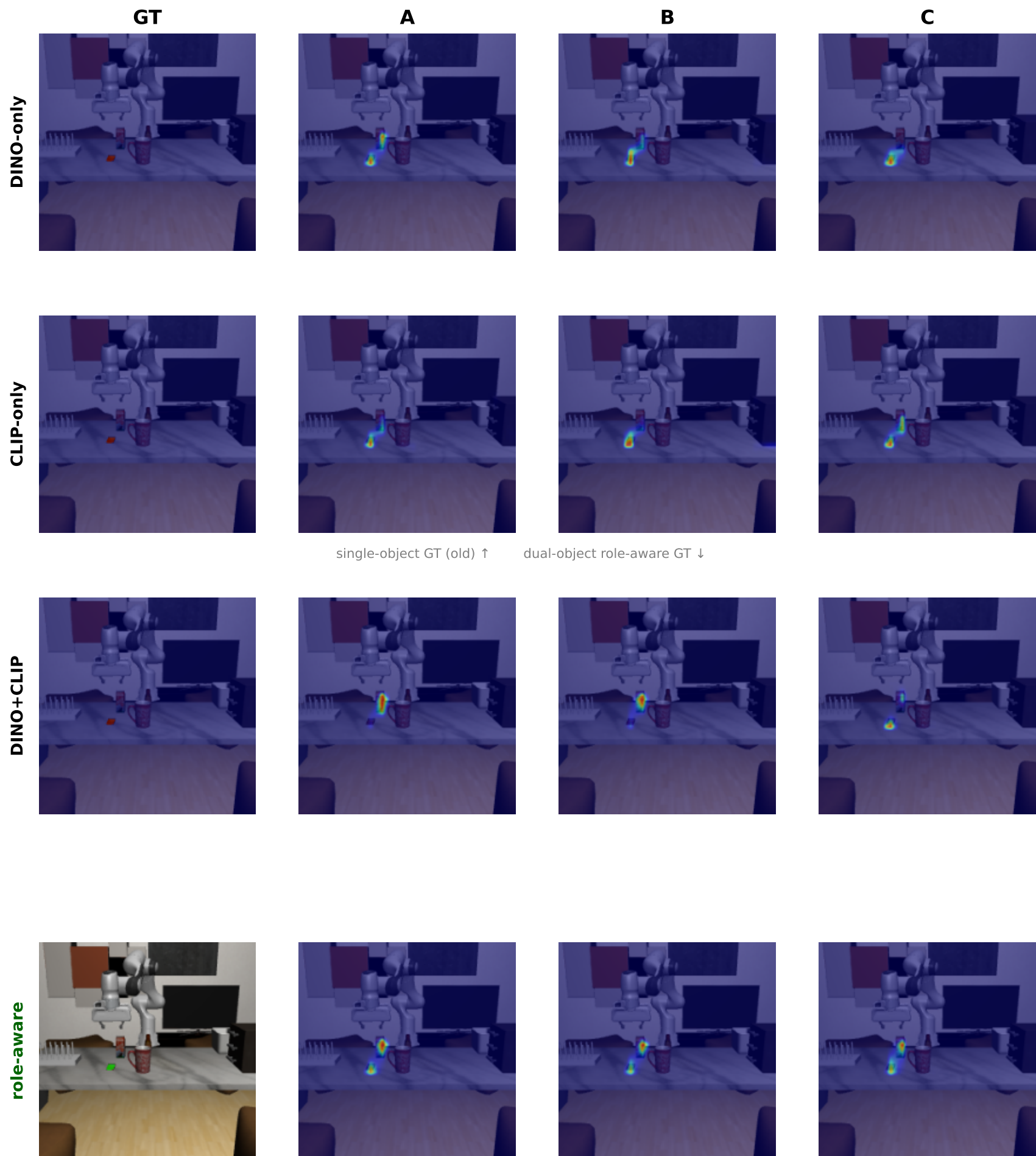


Mix_case7 — step: Pick up the butter from coffee table [grasp]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to



Mix_case7 — step: Place the butter onto the rack [release]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to



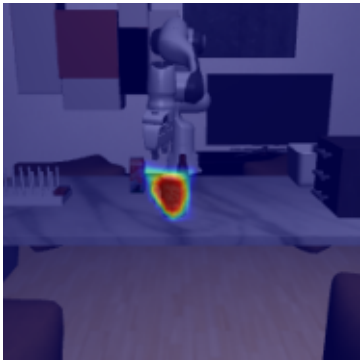
Mix_case7 — step: Pick up the red coffee mug from coffee table [grasp]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to

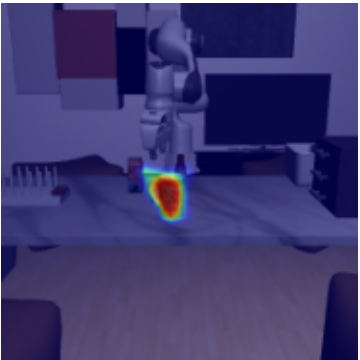
GT



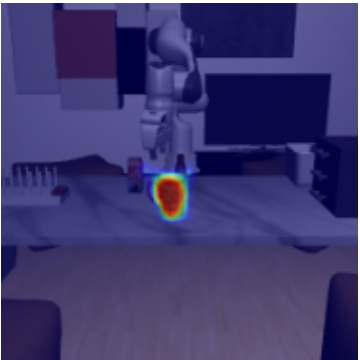
A



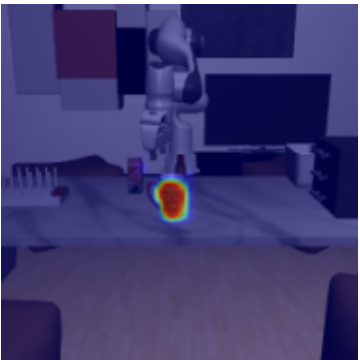
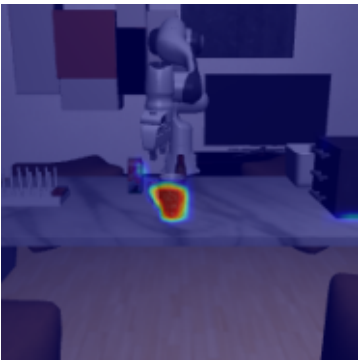
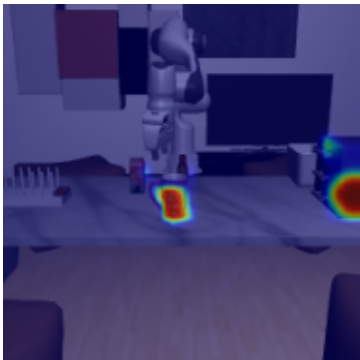
B



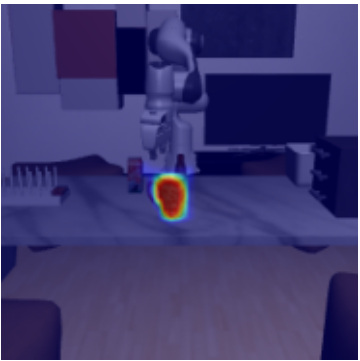
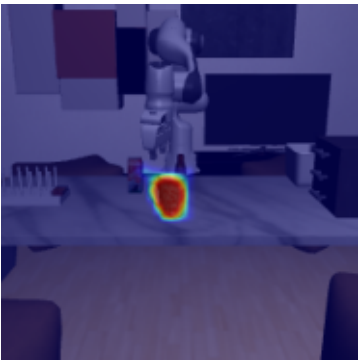
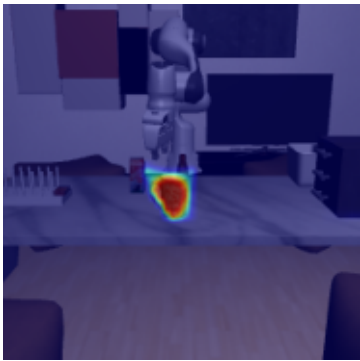
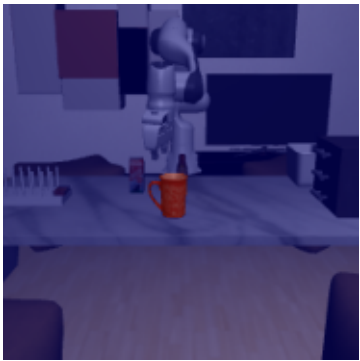
C



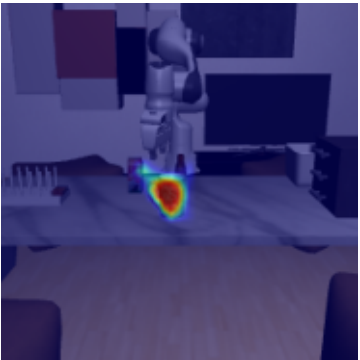
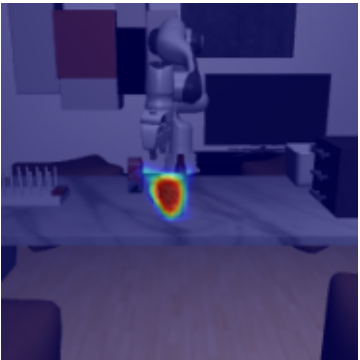
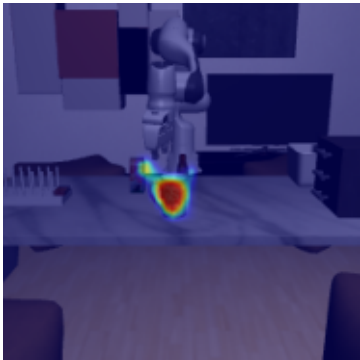
DINO-only



CLIP-only



DINO+CLIP



role-aware

single-object GT (old) ↑ dual-object role-aware GT ↓

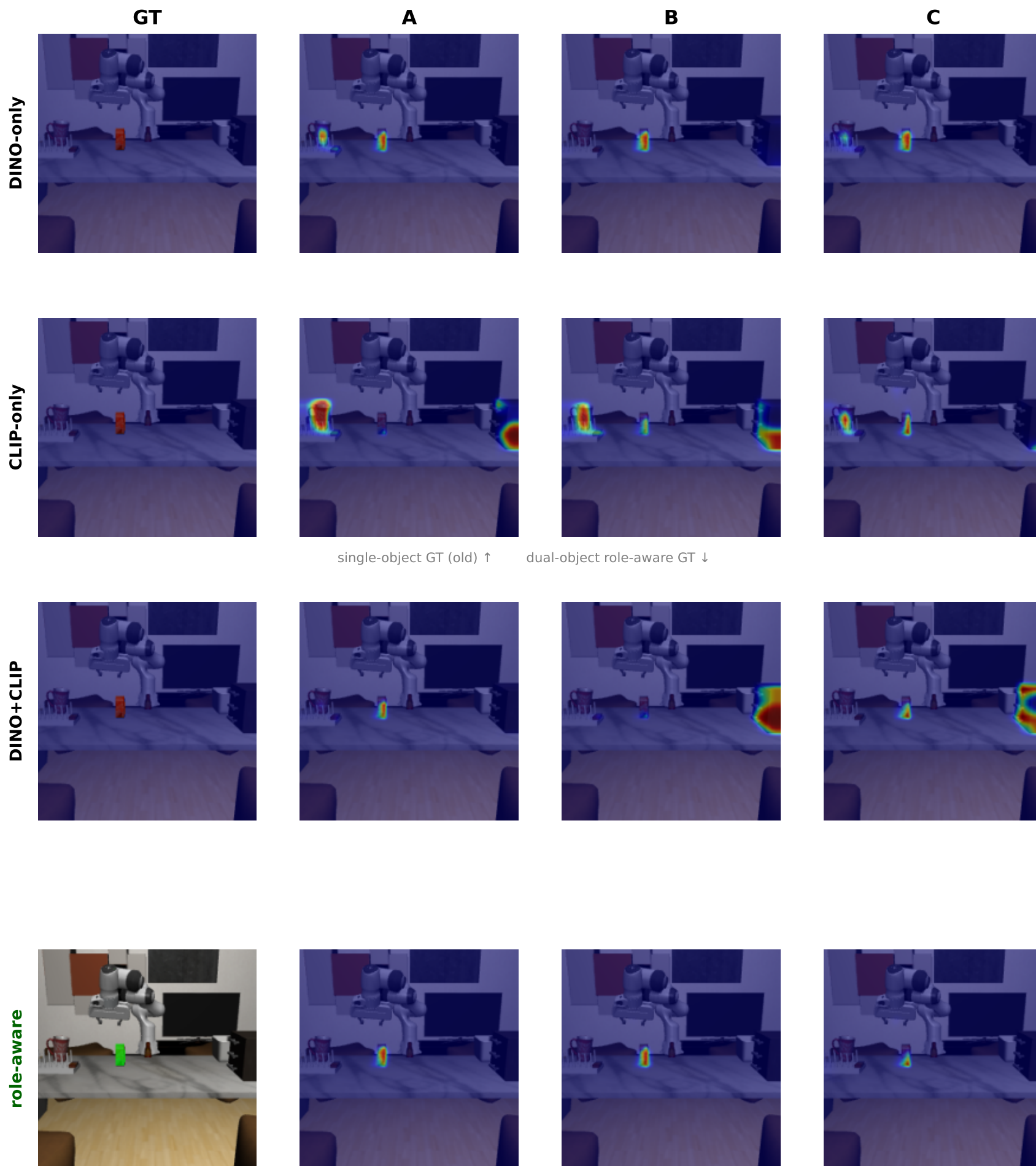
Mix_case7 — step: Place the red coffee mug onto the rack [release]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to



Mix_case7 — step: Pick up the milk from coffee table [grasp]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to



Mix_case7 — step: Pour out the milk from the milk container [manipulate]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to

GT



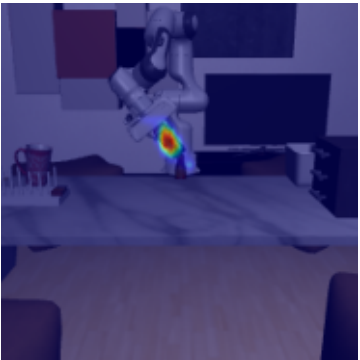
A



B



C



DINO-only



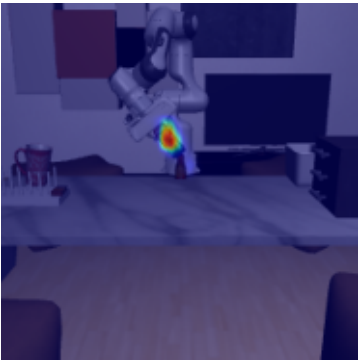
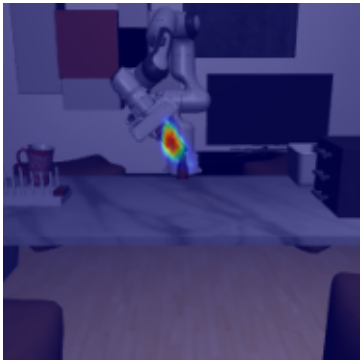
CLIP-only



DINO+CLIP

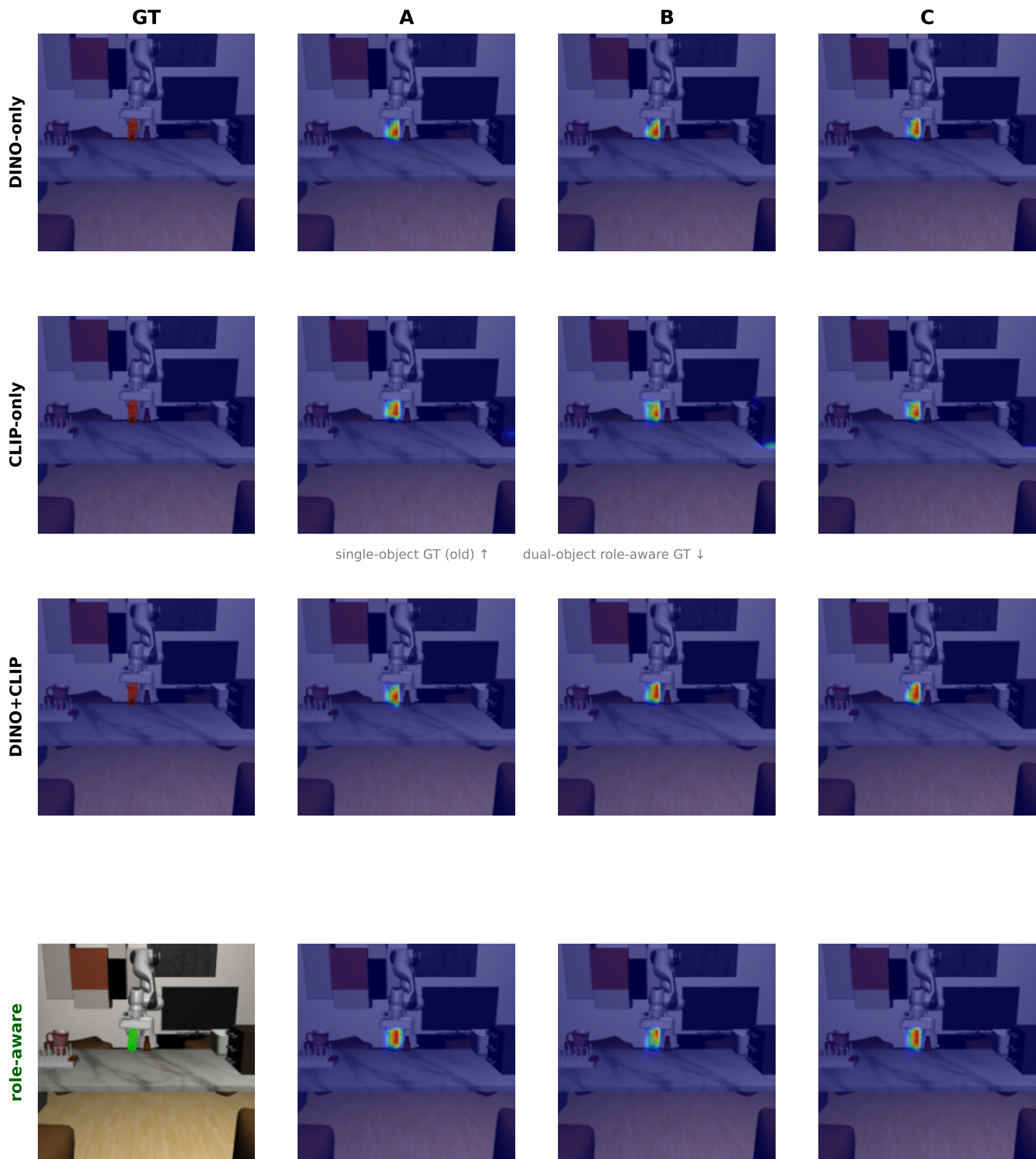
single-object GT (old) ↑ dual-object role-aware GT ↓

role-aware



Mix_case7 — step: Place down the milk onto coffee table [release]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to



Mix_case7 — step: Open the wooden cabinet top region [manipulate]

goal: After putting the butter and mug on the rack, pour in the milk, and check the wooden cabinet to

